*Neha Moolchandani Course: Data Mining | Professor: Dr. Chengcui Zhang*

I Neha Moolchandani declare that I have completed this assignment completely and entirely on my own, without any consultation with others.  I understand that any breach of the UAB Academic Honor Code may result in severe penalties.

*Assign 4*

Records

**Assignment #4:**

1. This problem uses the data table of Table 8.1 of the text (p. 338). Consider the first 5 records as the entirety of the table. (The main purpose of part (a) of this problem is to explain the relation between the file of “items” that Apriori and FP-growth work on, and data files such as this one.)

Interpreting the records of a file as sets of items: the set of all items is the union of the sets of values appearing in the attributes.

Graphical user interface, table

Description automatically generated

Note 1: RecordID is not to be considered an attribute.

Note 2: Because columns 3 and 5 have values in common, recode values in column 3 as “no3” and “yes3” and values in column 5 as “no5” and “yes5”.

Thus the set of items is {youth, mid-aged, senior; high, … ; no5, yes5}.

1. The first record, as a set of items, is {youth, high, no3, fair, no5}. Write the remaining 4 records of the (truncated) file as sets of items. (16 pts)
2. {youth, high, no3, fair, no5}
3. {youth, high, no3, excellent, no5}
4. {middle\_aged, high, no3, fair, yes5}
5. {senior, medium, no3, fair, yes5}
6. {senior, low, yes3, fair, yes5}
7. Letting the minimum support be 3 records, find F1, C2, F2, C3, and F3 (or, using the notation of pages 249-253 of the text, L1;C2;L2;C3;L3). (20 pts)

Text

Description automatically generated

Calendar

Description automatically generated

1. {youth, high, no3, fair, no5}
2. {youth, high, no3, excellent, no5}
3. {middle\_aged, high, no3, fair, yes5}
4. {senior, medium, no3, fair, yes5}
5. {senior, low, yes3, fair, yes5}

Support = Freq(a.b) / N

MinSup = 3 C1 -> 1st Scan:

|  |  |
| --- | --- |
| **ItemSet** | **Support** |
| {youth} | 2 |
| {middle\_aged} | 1 |
| {senior} | 2 |
| {High} | 3 |
| {Medium} | 1 |
| {Low} | 1 |
| {Fair} | 4 |
| {No3} | 4 |
| {yes3} | 1 |
| {No5} | 2 |
| {Yes5} | 3 |

F1 -> Delete the 1 and 2 since MinSUpport is 3 so delete Youth, MiddleAged, Senior, Medium, Low, Yes3 and No5

Left With:

|  |  |
| --- | --- |
| **ItemSet** | **Support** |
| {High} | 3 |
| {Fair} | 4 |
| {No3} | 4 |
| {Yes5} | 3 |

C2:

|  |
| --- |
| **ItemSet** |
| {High,Fair} |
| {High,No3} |
| {High,Yes5} |
| {Fair,No3} |
| {Fair, Yes5} |
| {No3, Yes5} |

C2: Second Scan

|  |  |
| --- | --- |
| **ItemSet** | **Support** |
| {High,Fair} | 2 |
| {High,No3} | 3 |
| {High,Yes5} | 1 |
| {Fair,No3} | 3 |
| {Fair, Yes5} | 3 |
| {No3, Yes5} | 2 |

F2 -> Delete the 1 and 2 since MinSUpport is 3 so delete Row1, Row3, and Row6

Left With:

|  |  |
| --- | --- |
| **ItemSet** | **Support** |
| {High,No3} | 3 |
| {Fair,No3} | 3 |
| {Fair, Yes5} | 3 |

C3:

|  |
| --- |
| **ItemSet** |
| {Fair,No3} |

F3: 3rd Scan:

|  |  |
| --- | --- |
| {Fair,No3} | 3 |

2. (a) construct the FP-tree for the set of records below, using minimum support threshold 1. This tree is denoted as *T*. (Items are already in order by decreasing support.) (12 pts)

{}

a: 5

b:5

c:5

d:4

e:4

f:1

g: 1

Root(Null)

/ \

A:4 B:3

/ \ \

B:2 C:2 D:2

/ \ \

C:2 E:2 G:1

/ \

D:1 F:1

/

E:1

1. Construct the conditional FP-tree for item *f*, which will be denoted as *Tf*. (12 pts)

Item Conditional DataBase

A empty

B empty

C ab:2, a:2

D abc:1, ac:1, b:2

E abc:1, ac:2

F ace:1

G bd:1

1. Execute the procedure FP-growth(*Tf*, ), where ={*f*}. Show the step by step details and the generated patterns together with their support counts. (p. 260) (12 pts)

Selected and sorted in order of L: a: 5, b:5, c:5, d:4, e:4, f:1, g:1

Root(Null)

/ \

A:4 B:3

/ \ \

B:2 C:2 D:2

/ \ \

C:2 E:2 G:1

/ \

D:1 F:1

/

E:1

1. Construct the conditional FP-tree for item *d*, which will be denoted as *Td*. (10 pts)

Item Conditional DataBase

A empty

B empty

C a:4

D empty

E a:3, c:3

F a:1, c:1, 3:1

G b:1, d:1

1. Execute FP-growth (*Td*, {*d*}). Show the step by step details and the generated patterns together with their support counts. (18 pts)

Removing F and G as MinSUpport is 1

Constructing the Tree based off

The set of items is I = {a, b, c, d, e, f, g}

a: 5, b:5, c:5, d:4, e:4, f:1, g:1

F-list: a,b,c,d,e,g

The set of records is

below.

LIST

ORDERED FREQ ITEMLIST

A,b,c,d,e

A,c,e,f

B,d,g

a,b,c,e

a,c,e

b,d

b

a,c,d

a, b, c, d, e

a, c, e, f

b, d, g

a, b, c, e

a, c, e

b, d

b

a, c, d